

Machine Learning & Data Mining

Prova intermedia 20 aprile 2018

1) Concept learning

1.a) Which is the sequence of hypotheses generated by the execution of algorithm FIND-S for learning the “EnjoySport” concept using the following training instances?

Sky	Temp	Humid	Wind	Water	Forecast	EnjoySpt
Sunny	Warm	Normal	Strong	Warm	Same	Yes
Sunny	Warm	High	Strong	Warm	Same	Yes
Rainy	Cold	High	Strong	Warm	Change	No
Sunny	Warm	High	Strong	Cool	Change	Yes

1.b) What is the meaning of UNbiased learner? What would be an unbiased learner for EnjoySport?

2) Decision trees

2.a) Describe the method of splitting based on the Classification Error.

2.b) If the leaf nodes of a decision tree correspond to subsets of training records that are not uniformly classified, how can we decide the classification of a new instance using the tree?

3) Valutazione di algoritmi

3.a) How is measured the accuracy of a learning algorithm for a training data set D using the K-fold cross validation method?

3.b) When is the case that the training data set can generate overfitting in a decision tree build using the general algorithm?

4) Reti Neurali

4.1) Describe what is a linear unit (unthresholded Perceptron) and which is the error function that we need to minimize for learning its weights.

4.2) Describe the meaning of the following formula for updating the weights of a feed-forward neural network at iteration n of the backpropagation algorithm:

$$\Delta w_{ji}(n) = \eta \delta_j x_{ji} + \alpha \Delta w_{ji}(n-1)$$

5) Learning Bayesiano

5.1) Which are the probabilistic assumptions made by the of Naive Bayes Classifier? Why in practice often it works well for classification even if the probabilistic values it computes are only approximations?

5.2) How is the Optimal Bayes Classifier formally defined. Why in practice it is difficult to implement it?